Physica A 525 (2019) 290–303

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Analysis of daily streamflow complexity by Kolmogorov measures and Lyapunov exponent

Dragutin T. Mihailović [a], Emilija Nikolić-Đorić [a], Ilija Arsenić [a], Slavica Malinović-Milićević [b], Vijay P. Singh [c], Tatijana Stošić [d], Borko Stošić [d,*]

[a] Faculty of Agriculture, University of Novi Sad, Dositej Obradovic Sq. 8, 21000 Novi Sad, Serbia
[b] ACIMSI - Center for Meteorology and Environmental Modelling, University of Novi Sad, Dositej Obradovic Sq. 7, 21000 Novi Sad, Serbia
[c] Department of Biological and Agricultural Engineering and Zachry Department of Civil Engineering, Texas A and M University, College Station, TX 77843-2117, USA
[d] Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco, Rua Dom Manoel de Medeiros s/n, Dois Irmãos, 52171-900, Recife, Brazil

HIGHLIGHTS

- Daily streamflow complexity is analyzed using Kolmogorov complexity measures.
- Highest complexity discrete spectrum indicates the highest streamflow randomness.
- Lyapunov exponent and time of the streamflow series predictability are also studied.
- The corrected Lyapunov time is inversely proportional to the Kolmogorov complexity.

ARTICLE INFO

Article history:
Received 29 August 2018
Received in revised form 28 December 2018
Available online 20 March 2019

Keywords:
Streamflow time series
Kolmogorov complexity based measures
Lyapunov exponent
Lyapunov time
Kolmogorov time
Predictability of streamflow time series

ABSTRACT

Analysis of daily streamflow variability in space and time is important for water resources planning, development, and management. The natural variability of streamflow is being complicated by anthropogenic influences and climate change, which may introduce additional complexity into streamflow records. To address the complexity in streamflow, daily discharge data recorded during the period 1989–2016 at twelve gauging stations on Brazos River in Texas (USA) were used to derive a set of novel quantitative tools: Kolmogorov complexity (KC) and its derivative-associated measures to assess complexity, and Lyapunov time (LT) to assess predictability. It was found that all daily discharge series exhibited long memory with an increasing down-flow tendency, while the randomness of the series at individual sites could not be definitively concluded. All Kolmogorov complexity measures had relatively small values with the exception of the USGS (United States Geological Survey) 08088610 station at Graford, Texas, which exhibited the highest values of the complexity measures. This finding may be attributed to the elevated effect of human activities at Graford, and proportionally lesser effect at other stations. In addition, complexity tended to decrease downflow, meaning that larger catchments were generally less influenced by anthropogenic activities. The correction on randomness of Lyapunov time (quantifying predictability) was found to be inversely proportional to the Kolmogorov complexity, which strengthened

* Corresponding author.
E-mail addresses: guto@polj.uns.ac.rs (D.T. Mihailović), emily@polj.uns.ac.rs (E. Nikolić-Đorić), ilija@polj.uns.ac.rs (I. Arsenić), slawica@sbb.rs (S. Malinović-Milićević), vsingh@tamu.edu (V.P. Singh), tastosic@gmail.com (T. Stošić), borkostosic@gmail.com (B. Stošić).

https://doi.org/10.1016/j.physa.2019.03.041
0378-4371/© 2019 Elsevier B.V. All rights reserved.

our conclusion regarding the effect of anthropogenic activities, considering that KC and LT were distinct measures, based on rather different techniques.
© 2019 Elsevier B.V. All rights reserved.

1. Introduction

Streamflow varies in space and time and investigation of its variability is essential for planning, operation, and management of water resources systems. It is significantly influenced by human activities and climatic change, and represents a complex system. In the past two decades, several information measures and methods of nonlinear dynamics have been employed to explain the complex and chaotic nature of streamflow [1–9]. For example, some practical guidelines for an appropriate basis and a suitable methodology towards a classification framework in hydrology have been offered by [10]. They pointed out the importance of nonlinearity and complexity in hydrologic systems. The nonlinearity and complexity of complex systems are also important in the study of reliability of hydrological models. The question arises: What is nonlinearity? Nonlinearity is a relationship which cannot be explained as a linear combination of its variable inputs. In other words, the output does not change in proportion to a change in any of the inputs. However, it should be mentioned that in applying nonlinearity to earth surface systems and correspondingly to the hydrological systems, the terms "input" and "output" may need to be more generally and flexibly defined than to their mathematical or systems theory. Definition of complexity given by [11]: "Complexity characterizes the behavior of a system or models whose components interact in multiple ways and follow local rules, meaning there is no reasonable higher instruction to define the various possible interactions" is maybe literally the clearest but without unique quantification. Namely, complexity is weakened by the lack of an unambiguous definition. Such definitions abound for both dynamical systems and biological organisms but have drawbacks of either a conceptual or practical nature. Nonlinearity is associated with various aspects of nonlinear behavior, such as chaos, fractals and complex self-organization [12]. While this amalgam of nonlinearity and complexity may be regrettable and at least partly inaccurate, the connotations are deep-rooted. Note that not all nonlinear dynamics are complex and similarly not all complexity is nonlinear or intrinsic to system behavior (for example, small values of the Lyapunov exponent and high complexity of stochastic chaos). However, in the intersection of the nonlinearity and complexity sets, we encounter Complex Nonlinear Dynamics, which represents the complexity that does arise from intrinsic nonlinearities.

One of the information measures for quantifying randomness of streamflow is the Kolmogorov complexity (KC). This measure and its derivatives — Kolmogorov spectrum, Kolmogorov complexity spectrum highest value (KCM) [10], running Kolmogorov complexity (RKC), and mean Kolmogorov complexity (MKC) - provide information complementary to that of entropy, offering additional insight into the behavior of complex systems. However, Kolmogorov complexity (KC) has rarely been used to analyze streamflow time series, although it is capable of providing supplementary description (to that of entropy) on the complexity or degree of randomness in highly complex systems. Some attempts to use KC and its derivatives for analyzing streamflow were made by [6,8,13]. The KC and its derivatives are important to use, because randomness is a key characteristic that has important implications for streamflow modeling and prediction [14,15]. Another complementary measure important for streamflow analysis is the largest Lyapunov Exponent (LE), because it can detect the presence of deterministic chaos in complex hydrological systems and quantify the predictability of potential future outcomes [16–18].

The objective of this paper therefore is to investigate the complexity, chaotic behavior, and predictability of daily streamflow of Brazos River in Texas (USA) for the period 1989–2016, using KC, KC spectrum, KCM, MKC, RKC, and LE, since it is found that the shorter the time scale, the stronger the nonlinearity. All annual series are linear; whereas all daily streamflow processes are strongly nonlinear [19]. These analyses reveal those aspects of the phenomenon that are complementary to those of the classical statistical approach. More precisely, the empirical study performed in the current work (that should not be confused with an attempt to modeling), sheds new light on the general properties of the observed time series (intrinsic randomness and predictability). These aspects should in turn be taken into account when devising new heuristic or mathematical models, but can also be useful for detecting "unusual" behavior at some selected locations, such as at the USGS station 08088610 (Brazos River near Graford, TX), where high KC complexity may be attributed to thereby Morris Sheppard Hydroelectric Power plant.

The paper is organized as follows. Section 2 describes the Kolmogorov complexity and information measures derived from it (Kolmogorov complexity spectrum and its highest value, and running and mean Kolmogorov complexity), Lyapunov exponent, and Lyapunov time. Section 3 provides information on streamflow data and gauging locations. Finally, Section 4 presents the results obtained and discusses the nonlinear feature of streamflow time series, Kolmogorov complexity measures and complexity spectrum, and Lyapunov exponent, and the predictability of daily discharge. The concluding remarks are given in Section 5.

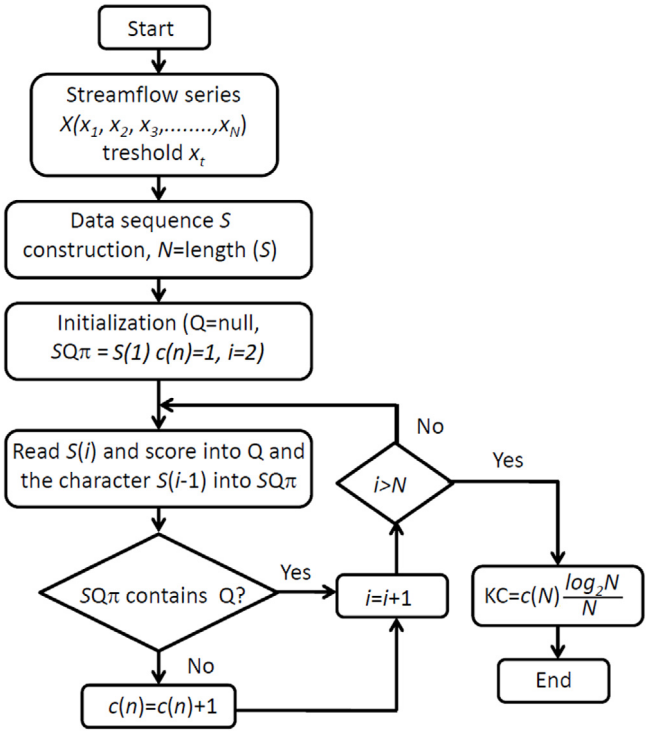

Fig. 1. Flowchart for calculation of the Kolmogorov complexity (KC) using the Lempel–Zev algorithm (LZA) [27].

2. Methods

2.1. Kolmogorov complexity and its derivatives

2.1.1. Kolmogorov complexity

The Kolmogorov complexity [20] stems from algorithmic information theory, which has been increasingly employed over the past 15 years for characterization of complex phenomena [6,8,13,21,22]. Let X denote streamflow and x its specific value. Kolmogorov complexity $K(x)$ of an object x is the length, in bits, of the smallest program that can be run on a universal Turing machine U and that prints object x (a *Turing machine* is a mathematical model of an abstract *machine*, which can use a predefined set of rules to determine a result from a set of input variables). While the complexity $K(x)$ is not directly computable for an arbitrary object x, in practice it is approximated by the size of the ultimate compressed version of x [23,24]. More precisely, a binary object x is compressed, and the size of the compressed object is identified with Kolmogorov complexity $K(x)$. We use here the Lempel and Ziv Algorithm (LZA) proposed [23] for calculating KC of a time series. In the following we shortly describe the calculation of the KC complexity of a time series $X\,(x_1, x_2, x_3, \ldots, x_N)$ by the LZA algorithm. It includes the following steps. (1) Encoding the time series by creating a sequence S of the characters 0 and 1 written as $s\,(i)\,, i = 1, 2, \ldots, N$, according to the rule $s\,(i) = 0$ if $x_i < x_t$ or 1 if $x_i > x_t$, where x_t is a threshold. The threshold is commonly selected as the mean value of the time series [21], while other encoding schemes are also available [25]; (2) calculating the complexity counter $c(N)$, defined as the minimum number of distinct patterns contained in a given character sequence. The complexity counter $c\,(N)$ is a function of the sequence length N, bounded by $b\,(N) = N/\log_2 N$ as N approaches infinity, i.e. $c\,(N) = O\,(b\,(N))$; (3) calculating the normalized information measure $C_k\,(N)$, defined as $C_k\,(N) \equiv c\,(N)\,/b\,(N) = c\,(N)\,\frac{\log_2 N}{N}$. For a nonlinear time series, $C_k\,(N)$ varies between 0 and 1, although it can be larger than 1 for random finite size sequences [22]. The LZA algorithm is described in condensed form [23,26]. Note that the pattern is a sequence in the coded time series which is unique and non-repeatable. A flow chart for the calculation of KC of a streamflow series $X\,(x_1, x_2, x_3, \ldots, x_N)$ using the LZA algorithm is shown in Fig. 1.

2.1.2. Kolmogorov complexity spectrum and its highest value

The Kolmogorov complexity of streamflow has two weaknesses: (i) It cannot distinguish between time series with different amplitude variations and that with similar random components; and (ii) in the conversion of a time series into a binary string, its complexity is unseen in the rules of the applied procedure. Therefore, in defining a threshold for a criterion for coding, some information about the composition of time series could be lost. In streamflow analysis, two measures were used: (i) Kolmogorov complexity spectrum (KC spectrum) and (ii) the highest value of KC spectrum,

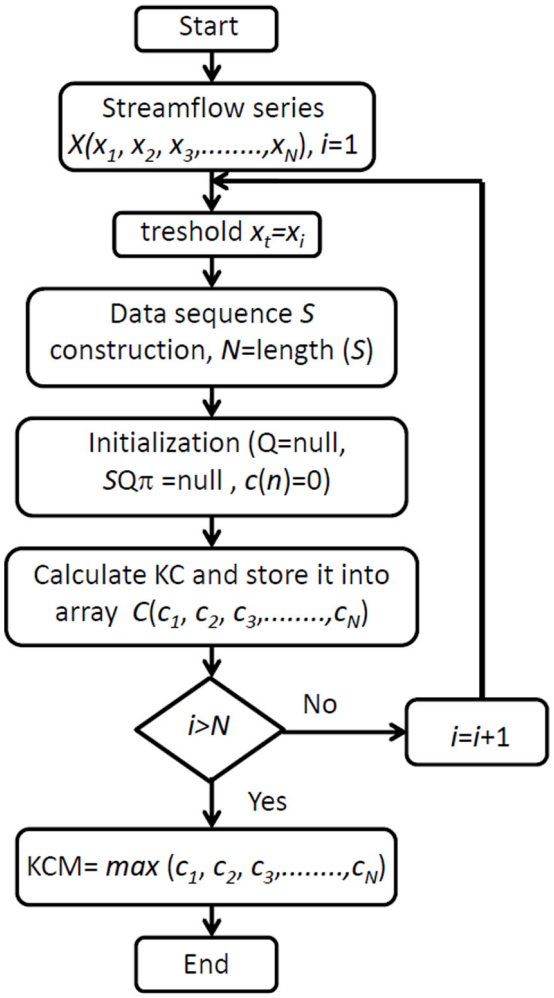

Fig. 2. Flow chart for calculation of the Kolmogorov complexity spectrum and its highest value (KCM) [27].

introduced by [13] who described the procedure for calculating the KC spectrum. Fig. 2 schematically shows how to calculate the KC spectrum $C(c_1, c_2, c_3, \ldots, c_N)$ for streamflow series $X(x_1, x_2, x_3, \ldots, x_N)$. This spectrum allows to investigate the range of amplitudes in a time series that represents a complex system with highly enhanced stochastic components. It may be noted that for a large number of samples of a time series, the computation of KC spectrum can be computationally challenging. Therefore, it is reasonable to divide the domain, including all values between the minimum (c_{min}) and maximum (c_{max}), into subintervals, which then are used as thresholds. In this study, 500 thresholds were used to obtain all spectra. The highest value K_m^C as in this series, i.e., $K_m^C = max\{c_i\}$, is the highest value of Kolmogorov complexity spectrum (KCM), as seen in Fig. 2.

2.1.3. Mean Kolmogorov complexity

Using the mean-value theorem and Simpson's rule, the mean Kolmogorov complexity K_C^M (MKC) can be calculated as

$$K_C^M = \frac{\Delta c}{3\,(c_n - c_0)}\,(c_0 + 4c_1 + 2c_2 + 4c_3 + 2c_4 + \cdots + 4c_{n-1} + 4c_n)\,, \tag{1}$$

here $c_0 = min\{c_i\}$ and $c_n = max\{c_i\}$, while the interval $[c_0, c_n]$ is divided into an even number n of subintervals of equal length $\Delta c = (c_n - c_0)/n$ using the $n + 1$ points.

2.1.4. Running Kolmogorov complexity

In statistical hydrology, a running mean [28] is often calculated to analyze data by creating series of averages of different subsets of the full data set that is a type of finite impulse response filter. This procedure was used to calculate the running Kolmogorov complexity (RKC). For a given series of streamflow values we extracted a fixed window (of size

1095 in this study) as in running mean procedure, and then the procedure for calculating the KC (Fig. 1) was applied. After that, the window was moved forward for a step with size of one day and the KC algorithm was applied, until the end of time streamflow series was reached.

2.2. Lyapunov exponent and Lyapunov time of streamflow time series

The Lyapunov exponent of a dynamical system is a quantity that characterizes the rate of separation of infinitesimally close trajectories [29] in parameter space. Positive Lyapunov exponent (LE) indicates that small fluctuations can lead to drastically different system behavior (small differences in the initial state lead to large differences in a later state). The term "butterfly effect" was coined by Edward Lorenz to describe this effect in a metaphorical example where flapping of the wings of a butterfly may lead to a faraway tornado in a distant future.

Quantitatively, two trajectories in phase with initial separation δX diverge at a rate given by $|\delta X(t)| \approx e^{\lambda t} |\delta X(0)|$ where λ is the Lyapunov exponent. It should be noted that the divergence is treated with the linear approximation. Because the rate of separation can be different for different orientations of the initial separation vector, there is a spectrum of Lyapunov exponents whose largest value is commonly the LE. A positive value of this exponent is usually taken as an indication that the system is chaotic. In this study, we obtained the LE for the standardized daily discharge time series by applying the Rosenstein algorithm [30], which was implemented in MATLAB program [31]. In this algorithm λ was calculated as

$$\lambda = \lim_{\tau \to \infty} \lim_{\varepsilon \to 0} \frac{1}{\tau} ln \left(\frac{|x(\tau) - x_{\varepsilon}(\tau)|}{\varepsilon} \right), \tag{2}$$

where $|x(0) - x_{\varepsilon}(0)| = \varepsilon$.

This algorithm is fast, easy to apply, and robust to changes in the embedding dimension, reconstruction delay, length of time series, and noise level. The applied MATLAB program calculated the proper embedding dimension and reconstruction delay. The value of embedding dimension was selected by FNN (False Nearest Neighbors) method or symplectic geometry method in the case of high noisy data [32,33]. The proper time delay was the lag before the first decline of autocorrelation value below $1/e$. For data with nonlinear dependency time delay was obtained by applying mutual information criteria [34]. The Lyapunov exponent relates to the predictability of measured time series which includes deterministic chaos as an inherent component. Model predictability here is understood as the degree to which a correct prediction of a system's state can be made either qualitatively or quantitatively. Causal determinism has a strong relationship with predictability in that perfect predictability requires strict determinism. However, lack of predictability does not necessarily mean lack of determinism, where limitations on predictability could be caused by factors such as a lack of information or high level of randomness. There exists no general rule to predict the time evolution of systems far from equilibrium, e.g. chaotic systems that do not approach an equilibrium state. Their predictability usually becomes progressively worse with time, and to quantify predictability, the rate of divergence of system trajectories in phase state can be measured (for example, Kolmogorov–Sinai entropy, and Lyapunov exponents [29]). In stochastic analysis a random process is considered predictable if it is possible to infer the next state from previous observations. In many models, however, randomness is a phenomenon which "spoils" predictability. Deterministic chaos does not mechanically denote total predictability but means that at least it improves the prognostic power. In contrast, stochastic trajectories cannot be projected into future. If $\lambda > 1$ then streamflow is not chaotic but is rather stochastic, and predictions cannot be based on chaos theory. However, if $0 < \lambda < 1$ it indicates an existence of chaos in streamflow. In that case, one can compute the approximate time (often called Lyapunov time) limit for which accurate prediction for a chaotic system is a function of LE. It designates a period, when a certain process (physical, mechanical, hydrological, quantum, or even biological) moves beyond the bounds of precise (or probabilistic) predictability and enters a chaotic mode. Thus, that time [35] can be calculated as

$$\Delta t_{lyap} = 1/\lambda. \tag{3}$$

If $\lambda \to 0$, implying that $\Delta t_{lyap} \to \infty$, then long-term accurate predictions are possible.

3. Data and computation

3.1. Data and gauging locations

Daily streamflow values were obtained from the National Water Information System: Web Interface at [36], for the Brazos River in Texas (USA) with a drainage area of approximately 118,000 km^2, extending from eastern New Mexico to more than 1000 km southeast to the Gulf of Mexico [37]. Daily streamflow observations from 12 USGS stream gauges on the mainstream were obtained for a period from 1989 to 2016, when simultaneous data for all the stations were available. The geographical locations of gauging stations are depicted in Fig. 3, and basic statistics of data are given in Table 1.

Daily discharge datawere standardized, i.e., for each calendar day "i" mean discharge $\langle x_i \rangle$ and standard deviation SD_i, over the year, "j" , were computed and then the standardized discharge on day "i" in year "j" was calculated as $y_{i,j} = \frac{(x_{i,j} - \langle x_i \rangle)}{SD_i}$ [9]. This procedure removes any seasonal effects.

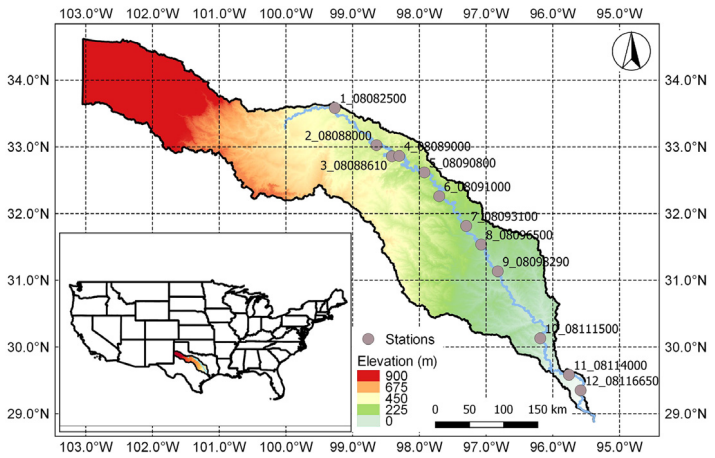

Fig. 3. Geographical locations of gauging stations on Brazos River used in this study [36].

Table 1
Basic descriptive statistics of daily discharge data of Brazos River for the period (1989–2016); (The first number indicates the order of the station used in this study).

USGS code	Station	Mean	Median	Min	Max	IQR	SD_I
1_08082500	Seymour	223.5	51.0	0.0	30 700.0	130.0	907.9
2_08088000	South Bend	613.0	110.0	0.0	43 800.0	320.0	2209.8
3_08088610	Graford	623.5	109.0	4.1	43 800.0	300.0	2306.9
4_08089000	Palo Pinto	723.7	133.0	8.5	39 700.0	361.0	2557.9
5_08090800	Dennis	974.4	195.0	0.0	79 500.0	418.0	3600.3
6_08091000	Glen Rose	1078.8	86.0	1.5	82 100.0	530.0	4093.9
7_08093100	Aquilla	1561.2	445.0	1.2	27 100.0	1118.0	3687.3
8_08096500	Waco	2456.1	695.0	0.5	44 000.0	1775.0	5237.7
9_08098290	Highbank	3103.7	873.5	30.0	70 300.0	2240.0	6148.1
10_08111500	Hempstead	8014.3	2520.0	58.0	137 000.0	7650.0	12 821.1
11_08114000	Richmond	8523.8	2855.0	182.0	102 000.0	8660.0	13 232.0
12_08116650	Rosharon	8851.4	3060.0	27.0	109 000.0	9080.0	13 638.0

3.2. Basic descriptive statistics

Basic descriptive statistics of daily discharge data of the gauging stations are summarized in Table 1, where for each station mean, median, minimum, maximum, interquartile range (IQR), and standard deviation (SD_i) are shown. It is seen from Table 1 that the differences between the maximum and the mean were in the range of roughly 10 to 40 standard deviations, strongly positively skewed, indicating a power law behavior. Indeed, frequency counts for the USGS 08082500 Brazos River station at Seymour, TX (USA), displayed in Fig. 4 on a log–log scale demonstrated a power law distribution, with similar behavior also observed at all other stations.

4. Results and discussion

4.1. Features of streamflow time series

Before analyzing streamflow complexity, flow characteristics were considered to establish potential dissimilarities between different streamflow series. The box and whisker plot is useful for comparing distributions of data sets [38]. Fig. 5 depicts a box and whisker plot of standardized daily discharge data and gives a clearer picture of the distribution of data in comparison with Table 1. It is seen that there was a smaller interval of variation at the third site, as well as a higher frequency of values smaller than the first quartile. The distributions of data for all stations had heavy right tails, stemming from a power law behavior.

All streamflow time series were characterized with a slow decay of positive r_k that is the estimate of the *kth* lag autocorrelation ρ_k. The values of lag k such that $|r_k| < 1/e$ presented in Table 2 indicated that the rate of autocorrelation decline was lower for sites (7–12).

If lag k increased, autocorrelations followed a power law, i.e., $r_k \sim ck^{\alpha}i + fk \to \infty$. Assuming self similarity of the stochastic process that generated time series, for $H > \frac{1}{2}$, the autocorrelation ρ_k behaved like $|\ k|^{\ 2-2H}$ [39]. Exponent α and Hurst exponent (H), as a measure of the long-term memory of a time series, were calculated using the function *hurstACVF(x)* of the R package *fractal* version 2.0-1 [40], from statistical program R 3.4.2. The method for estimating H

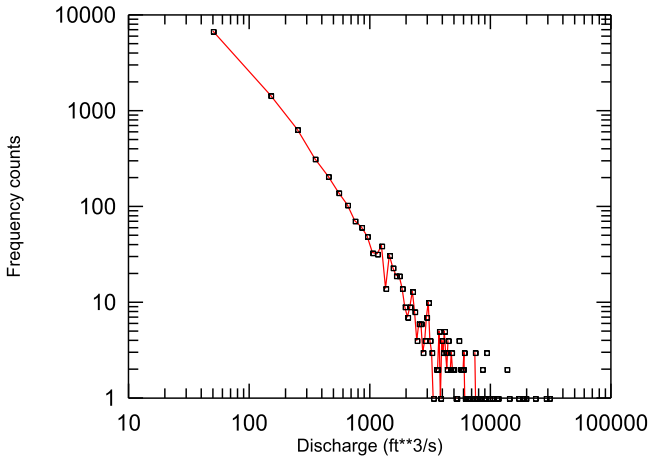

Fig. 4. Frequency counts of daily discharge data for the USGS 08082500 Brazos River station at Seymour, TX (USA) for the period (1989–2016).

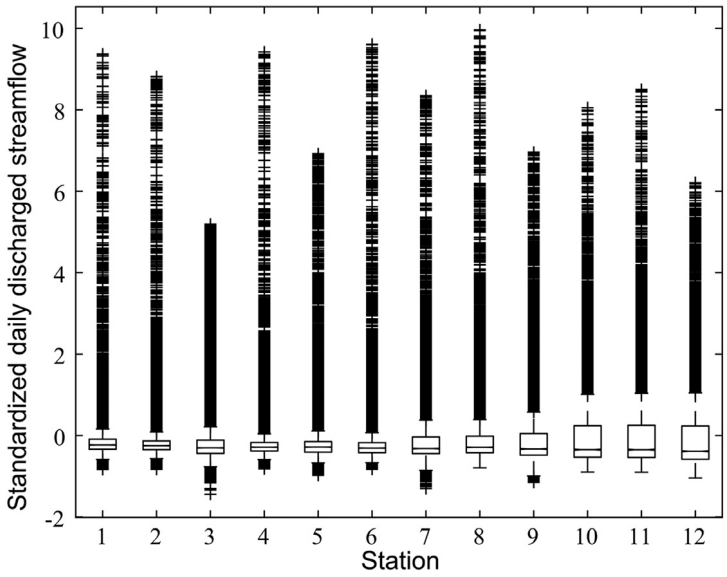

Fig. 5. Box and whisker plot of standardized daily discharge data on Brazos River.

Table 2
Values of lag k, exponent α and hurst exponent H for standardized daily discharge data on Brazos River.

USGS code	k	α	H	$H\,(R/S)$
1_08082500	13	−0.48	0.76	0.79
2_08088000	9	−0.47	0.76	0.81
3_08088610	10	−0.40	0.80	0.87
4_08089000	7	−0.48	0.76	0.81
5_08090800	9	−0.49	0.76	0.82
6_08091000	8	−0.48	0.76	0.81
7_08093100	32	−0.41	0.79	0.85
8_08096500	32	−0.43	0.79	0.84
9_08098290	31	−0.40	0.80	0.84
10_08111500	42	−0.36	0.82	0.83
11_08114000	43	−0.36	0.82	0.83
12_08116650	52	−0.33	0.84	0.84

was based on linear regression of scaled *asinh* of autocovariance function (ACVF) versus log (*lag*) over intermediate lag values. The values of H, that were between 0.5 and 1 (Table 2), indicated that the streamflow time series contained long

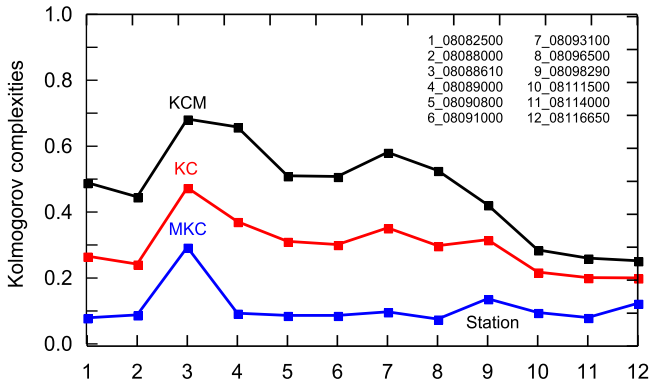

Fig. 6. Kolmogorov complexity (KC), the highest value of Kolmogorov complexity spectrum (KCM), and mean Kolmogorov complexity (MKC) of standardized daily discharge data on Brazos River.

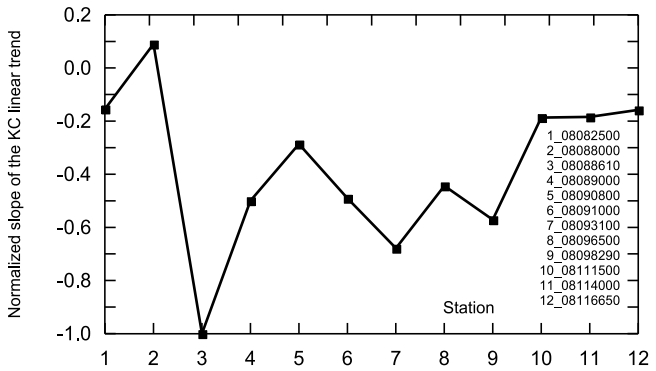

Fig. 7. Normalized slope of the KC linear trend of standardized daily discharge data of Brazos River.

memory or long range dependence [41]. This conclusion confirmed Hurst H (R/S) obtained by means of Rescaled Range Method using the function *hurstex(x)* of the *R*package*pracma.* Analysis showed that the standardized daily discharge time series exhibited long memory with a tendency to increase with the order of station except for site 3_08088610 where the series had a longer memory than did most stations. It should be noted that long range dependence may be spurious and be a consequence of nonstationarity, nonlinearity, and structural changes of streamflow [42]. Unlike random noise; long memory allows at least short-term predictability. Thus, one cannot conclude randomness of streamflow at individual stations.

4.2. Kolmogorov complexity measures and Kolmogorov complexity spectrum of standardized discharge daily data

The Kolmogorov complexity measures applied in this work shed additional light on the complex behavior of streamflow, complementing the usual entropy approaches commonly found in the literature. The values of KC, KCM, and MKC of standardized daily discharge data are shown in Fig. 6 which shows that the KC values for all daily streamflow time series were relatively small, ranging in the interval (0.20; 0.47). A similar behavior was observed for KCM and MKC, having values in the intervals (0.25; 0.68) and (0.08; 0.29), respectively. This is expected for lowland rivers in contrast to mountain rivers whose KC values can be up to 0.98 [8]. This feature of lowland rivers is clearly visualized through MKC (blue squares), which, on average, did not exceed 0.15. Although MKC offered some insight into streamflow complexity, it was still not sufficiently reliable because it was obtained as an average value from the Kolmogorov spectrum, indicating that some details had been lost. However, by KC (red squares) and KCM (black squares) more content information measures can be obtained. Between these two information measures, KC yielded less information about streamflow complexity than did KCM. That is, KC gives average information about the streamflow time series complexity, since its complexity remained hidden in the rules of the applied procedure. In contrast, KCM contained information about the highest complexity among all complexities observed in the KC spectrum, indicating the highest level of streamflow randomness. Thus, it can be said that MKC, KC and KCM were overall, average and highest measures, respectively, of the complexity of streamflow time series. Finally, Fig. 6 shows that all three measures had the highest value, indicating that the streamflow time series at site 3_08088610 had the highest complexity.

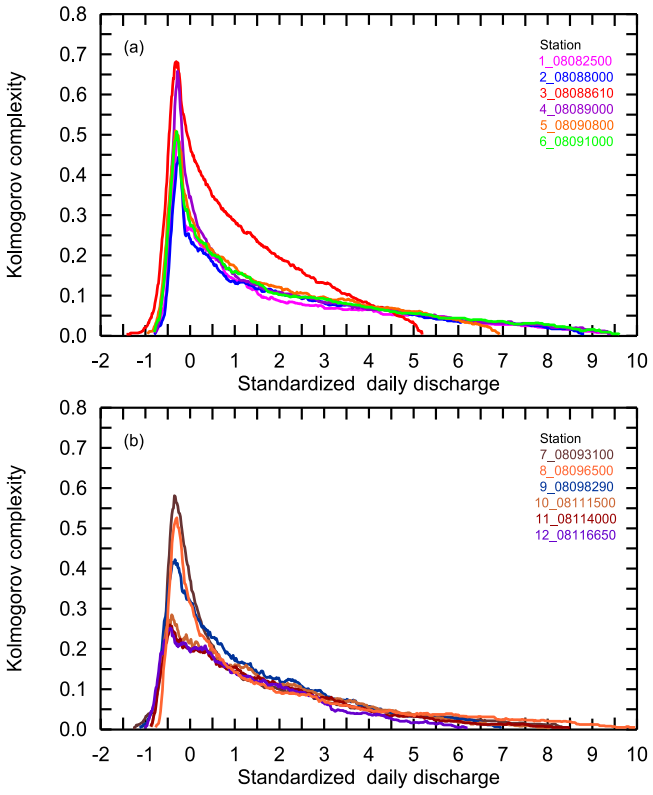

Fig. 8. Kolmogorov complexity spectra of standardized daily discharge data of Brazos River.

The above measures can be used to estimate the behavior of a long streamflow time series for (1989–2016, in this study). For example, a decreasing KC is a reliable indicator that the streamflow series becomes more uniform due to either environmental or human influences. First, RKC was applied to calculate the KC of daily streamflow data for all sites. Then, the slope of linear trend was obtained for each site and then all slopes were normalized by the highest absolute value. These values are depicted in Fig. 7 which shows that: (i) the normalized slopes of KC linear trends of streamflow time series for all the sites were negative (ranged between -1 and -0.15), except for site 2_0808800 whose normalized slope of the KC linear trend was positive (0.09); and (ii) site 3_08088610 had the highest negative value of the KC linear trend (-1). Apparently, there existed a systematic decrease of the KC along the Brazos River, except for site 2_0808800, likely due to the aforementioned reasons.

Fig. 8 shows the Kolmogorov complexity spectra of standardized daily discharge data. Inspection of Fig. 8a–b shows that the site with the highest KCM corresponded to the shortest range (Table 1) (site 3_08088610). In contrast, sites 4_8089000 and 7_08093100, also having high values of the KCM, covered a wider range. Quite naturally, a question arises: What does this spectrum offer? It gives information about the randomness of each amplitude in the streamflow time series using all information, i.e. all samples representing the time series.

Fig. 9 can be understood as a "streamflow discrete spectrum of the highest complexity" of standardized daily discharge data. This spectrum describes the presence of highest randomness (no other randomness left) in streamflow being influenced by many factors. This is a key element having important implications for modeling and prediction, since randomness is irreproducible and unpredictable [14,15].

In the above analysis we mentioned that change of KC complexity can be significantly induced by human activities. In that sense, our attention will be focused on the streamflow time series for site 3_08088610, since it has the highest values for all three KC complexity measures. There are many factors, both natural (runoff from rainfall and snowmelt; evaporation from soil and surface-water bodies; transpiration by vegetation; ground-water discharge from aquifers; ground-water recharge from surface-water bodies; sedimentation of lakes and wetlands; and formation or dissipation of glaciers, snowfields, and permafrost) and human-induced (surface-water withdrawals and transbasin diversions; river-flow regulation for hydropower and navigation; construction, removal, and sedimentation of reservoirs and storm water detention ponds; stream channelization and levee construction; drainage or restoration of wetlands; land-use changes such as urbanization that alter rates of erosion, infiltration, overland flow, or evapotranspiration; wastewater outfalls; and irrigation wastewater return flow), that cause continuous changes in streamflow time series and therefore in its complexity. All of the above factors can contribute to randomness either singly or in synergy with other factors. The

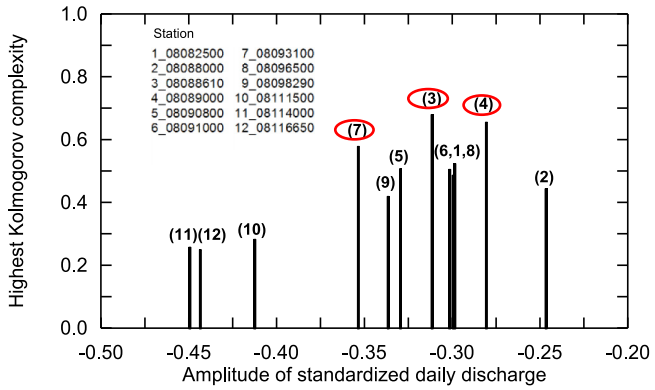

Fig. 9. The "streamflow discrete spectrum of the highest complexity" of standardized daily discharge data of Brazos River. The numbers in red ellipses correspond to the KC complexity spectrum highest values for three sites.

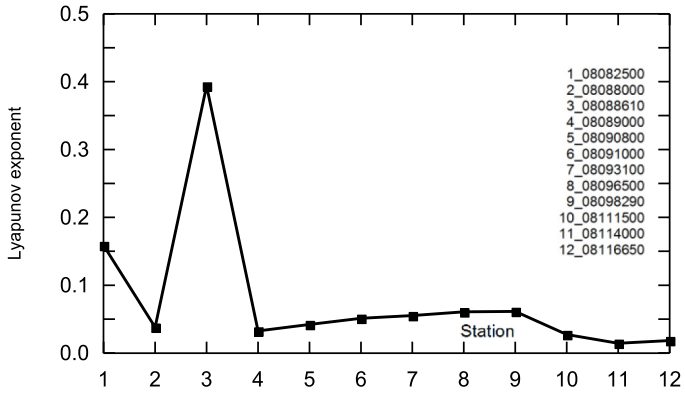

Fig. 10. Lyapunov exponent of standardized daily discharge data of Brazos River.

reason for the high KC complexity of this station may be attributed to the Morris Sheppard Hydroelectric Power plant at Morris Sheppard Dam (Possum Kingdom Reservoir) on the Brazos River in Palo Pinto County, built in the period 1938–1941 (11 miles southwest of Graford and 18 miles northeast from Mineral Wells). Currently, "USGS station 08088610 (Brazos River near Graford, TX) is located approximately 1.25 miles downstream of Possum Kingdom Reservoir. As such, this site is largely influenced by regulation. This gauge was established to monitor outflow from Possum Kingdom Reservoir. The gauge was initially located farther upstream, closer to the outflow from the reservoir. In 1995, the gage was moved downstream to the current location" (information obtained from [36]). Note that [8], using the KC complexity analysis, found that during 1946–1965, there was a change in complexity in the River Miljacka and the River Bosnia in comparison to the other chosen subintervals (analyzed time series was for the period 1926–1990). That complexity loss was interpreted as a result of intensive different human interventions on those rivers (establishing the network of channels and small dams for building the capacities for water consumption) after the Second World War.

4.3. Lyapunov exponent and predictability of standardized daily discharge data

Fig. 10 plots the Lyapunov exponent of standardized daily discharge data and shows that station 3_08088610 had the highest value of LE (0.3937), while all other sites had values in the interval (0.0183; 0.1585). In the last decade, there had been an increasing use of LE in daily streamflow analysis for detecting the deterministic chaos [16,17] and improving the chaotic approach to accurately predict potential future outgrowths. Using LE as an indicator, [16] investigated the potential existence of chaos in daily streamflow time series of Kizilirmak River (Iran). They established the presence of low chaos in the flow series with a positive value of LE (0.0061). Considering forecasting streamflow of Xijiang River (China) based on chaos radial basis function networks, [18] reported that LE was 0.1604. For all Brazos River sites, the LE values can be placed between the LE values of Kizilirmak River and Xijiang River. Only exception is site 3_08088610 having a much higher value. This feature as well as a longer memory than the most stations can be attributed to structural changes in streamflow resulting from human and environmental activities.

Using the LE method, one can obtain the Lyapunov time, Δt_{lyap} (LT) i.e. the time interval for making accurate predictions [18,43–45, among others]. Fig. 11a depicts the LE method for 12 sites on Brazos River given by the LT in days having the

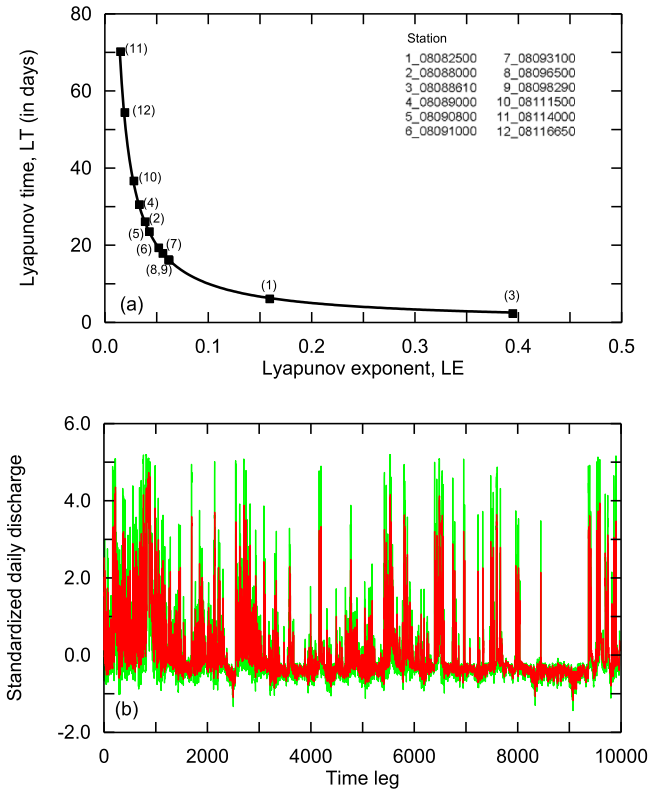

Fig. 11. (a) Predictability of standardized daily discharge data of Brazos River given by the LT (in days); (b) Standardized discharge daily streamflow for the site 3_08088610: measured (green) and simulated by the ARIMA model [46].

shape of a power function. Simple inspection indicates that the lowest predictability was for sites 3_08088610 (3 days) and 1_08082500 (6 days). Sites 8_08096500, 9_08098290, 7_08093100 and 6_08091000 had predictability of less than 20 days, while for sites 5_08090800, 2_08088000, 4_08089000 and 10_08114000, the prediction effect was between 20 and 40 days. Finally, the lowest LE values were for sites 11_08114000 and 12_08116650. Using this method [18] found that the prediction effect for the Xijiang River was poor (five-day prediction).

To illustrate the meaning and importance of the LT, which is not a predictability tool we applied an autoregressive integrated moving average (ARIMA (p, d, q)) model in its improved version [46] which is widely applied in hydrological simulations. This model has the property that uncertainty in a forecast increases exponentially with time. For the streamflow time series at site 3_08088610 the best ARIMA(5,1,3) model according to information criterion was identified, estimated, and used for prediction h-steps ahead (h = 1 in our simulation). The results of simulations are given in Fig. 11b. From this figure is seen that this model did not include the property of non-linearity which is a deterministic characteristic as well as a variable variance and the appearance of extreme values and leads to erroneous prediction. Note that LT for site 3_08088610 indicates that already after three days any conceivable model correctly representing the phenomenon should enter a chaotic mode and cannot capture the non-linearity of the process. The smallest predictability in the case of site 3_08088610 is not contradictory with long memory that characterizes this station. Long memory may be present in stochastic and deterministic processes. The ARIMA model is an example of stochastic long memory process. In that case the maximal Lyapunov coefficient is zero, and the H value may be high. On the other hand, in some special cases of chaotic time series, there is positive correlation between max LE and H [47]. H and max LE are complementary measures, as H cannot measure sensitivity to initial conditions. So the long memory does not imply high predictability if the model is sensitive to initial conditions.

The Lyapunov time was corrected for the presence of randomness, since the chaos theory generally deals with the "irregular behavior in a complex system that is generated by nonlinear deterministic interactions with only a few degrees of freedom, where noise or intrinsic randomness does not play an important role" [48]. The Lyapunov time was used for predicting the time series. However, many streamflow time series are highly complex. Therefore, Δt_{lyap} can be corrected for randomness in the following way. Similar to Δt_{lyap} we can introduce a randomness time $\Delta t_{rand} = \frac{1}{K_c}$ (in time units, second, hour or day), where K_c is the Kolmogorov complexity. Henceforth, we shall denote this quantity Kolmogorov time (KT), as it quantifies the time span beyond which randomness significantly influences predictability. Then, the Lyapunov time corrected for randomness is defined as $\left[0, \Delta t_{lyap}\right] \cap [0, \Delta t_{rand}]$. It could be stated that the KT designates the size of

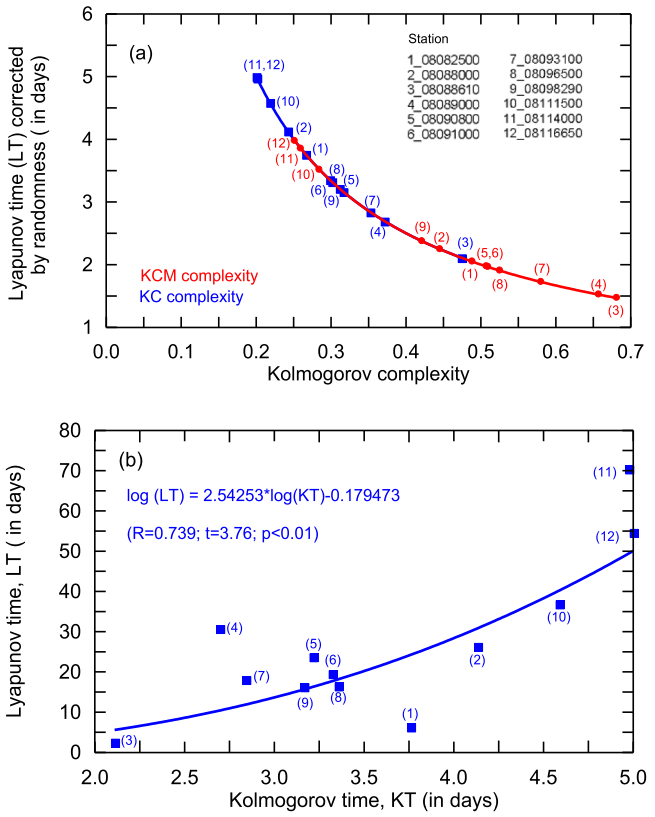

Fig. 12. (a) Predictability of standardized daily discharge data of Brazos River given by the Lyapunov time (LE) corrected by randomness (in days); (b) LT time versus KT time.

the time window within time series where complexity remains nearly unchanged. Fig. 12a shows the predictability of standardized daily discharge data given by the LT in days corrected for randomness. The figure shows that the prediction effect Δt_{rand} for KC is extended between two and five days, while for KCM it is between one and four days.

Fig. 12b depicts Lyapunov time (LT) plotted versus Kolmogorov time (KT). It is seen that the relationship between them is expressed through power law dependence. How much the randomness can reduce the LT can be seen from comparison of Figs. 11a and 12b. According to Fig. 11b the longest predictability in the LT units for the 12 sites at Brazos River was around 70 days, while following from the fit on Fig. 12b the presence of randomness reduced that number to 45 days.

5. Conclusions

Daily river flow data from twelve gauging stations on Brazos River in Texas (USA) was analyzed using the Kolmogorov complexity and related complexity measures (Kolmogorov complexity spectrum and its highest value, running, and mean Kolmogorov complexity), Lyapunov exponent and Lyapunov time. The following conclusions are drawn from this study:

(1) Standardized daily discharge time series exhibit long memory with a tendency to increase with the order of the gauging station with the exception of one station which has a longer memory than do most stations. Long range dependence may, however, be spurious and be the consequence of non-stationarity, nonlinearity and structural changes of time series. Unlike random noise, long memory allows for at least short-term predictability.
(2) The values of KC, KCM, and MKC of streamflow are relatively small, except for one site which has the highest value.
(3) All complexity measures can be used to estimate the behavior of streamflow time series for a long time period. Thus, a decrease of KC is a reliable indicator that streamflow series becomes more uniform due to environmental or human influences. The normalized slope of KC linear trend of streamflow series for all sites is negative, except for a site whose normalized slope is positive. There exists a systematic decrease of KC along the Brazos River, except for one site.
(4) Kolmogorov complexity spectra yield information about the randomness for each amplitude in the streamflow time series.
(5) The streamflow discrete spectrum of the highest complexity shows the presence of highest randomness in river flow, a key in modeling and prediction.

(6) The Kolmogorov time (KT) designates the size of the time window within time series where complexity remains nearly unchanged.

Acknowledgments

This paper was realized as part of the project "Studying climate change and its influence on the environment: impacts, adaptation and mitigation" (43007) financed by the Ministry of Education and Science of the Republic of Serbia within the framework of integrated and interdisciplinary research for the period 2011–2018. We also acknowledge support of Brazilian agencies CAPES and CNPq (grant No 310441/2015-3).

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] V.P. Singh, The use of entropy in hydrology and water resources, Hydrol. Process. 11 (1997) 587–626, http://dx.doi.org/10.1002/(SICI)1099-1085(199705).
[2] I. Krasovskaia, Entropy-based grouping of river flow regimes, J. Hydrol. 202 (1997) 173–191, http://dx.doi.org/10.1016/S0022-1694(97)00065-6.
[3] J.D. Salas, H.S. Kim, R. Eykholt, P. Burlando, T.R. Green, Aggregation and sampling in deterministic chaos: implications for chaos identification in hydrological processes, Nonlinear Proc. Geoph. 12 (4) (2005) 557–567, http://dx.doi.org/10.5194/npg-12-557-2005.
[4] M.S. Movahed, E. Hermanis, Fractal analysis of river flow fluctuations, Phys. A 387 (2008) 915–932, http://dx.doi.org/10.1016/j.physa.2007.10.007.
[5] S. Hajian, M.S. Movahed, Multifractaldetrended cross- correlation analysis of sunspot numbers and river flow fluctuations, Phys. A 389 (2010) 4942–4957, http://dx.doi.org/10.1016/j.physa.2010.06.025.
[6] S.K. Asok, Complexity analysis of riverflow time series, Stoch. Env. Res. Risk. 23 (2009) 361–366, http://dx.doi.org/10.1007/s00477-008-0222-x.
[7] B. Sivakumar, V.P. Singh, Hydrologic system complexity and nonlinear dynamic concepts for a catchment classification framework, Hydrol. Earth Syst. Sc. 16 (2012) 4119–4131, http://dx.doi.org/10.5194/hess-16-4119-2012.
[8] D.T. Mihailović, E. Nikolić-Đorić, N. Drešković, G. Mimić, Complexity analysis of the turbulent environmental fluid flow time series, Phys. A 395 (2014) 96–104, http://dx.doi.org/10.1016/j.physa.2013.09.062.
[9] T. Stosic, B. Stosic, V.P. Singh, Optimizing streamflow monitoring networks using joint permutation entropy, J. Hydrol. 552 (2017) 306–312, http://dx.doi.org/10.1016/j.jhydrol.2017.07.003.
[10] B. Sivakumar, V.P. Singh, Hydrologic system complexity and nonlinear dynamic concepts for a catchment classification framework, Hydrol. Earth Syst. Sc. 16 (2012) 4119–4131, http://dx.doi.org/10.5194/hess-16-4119-2012.
[11] S. Johnson, Emergence: The Connected Lives of Ants, Brains, Brains, Cities, 2001, p. 19, ISBN 3411040742.
[12] J.D. Phillips, Sources of nonlinearity and complexity in geomorphic systems, in: Progress in Physical Geography, Vol. 27, 2003, p. 123, http://dx.doi.org/10.1191/0309133303pp340ra.
[13] D.T. Mihailović, G. Mimić, E. Nikolić-Djorić, I. Arsenić, Novel measures based on the Kolmogorov complexity for use in complex system behavior studies and time series analysis, Open Phys. 13 (2015) 1–14, http://dx.doi.org/10.1515/phys-2015-0001.
[14] A.S. Khan, R. Ganguly, S. Saigal, Detection and predictive modeling of chaos in finite hydrological time series, Nonlinear Proc. Geoph. 12 (2005) 41–53, http://dx.doi.org/10.5194/npg-12-741-2005.
[15] B. Sivakumar, Chaos in Hydrology: Bridging Determinism and Stochasticity, first ed., Dordrecht, Netherlands, 2017.
[16] M.A. Ghorbani, O. Kisi, O.M. Alinezhad, A probe into the chaotic nature of daily streamflow time series by correlation dimension and largest Lyapunov methods, Appl. Math. Model. 34 (2010) 4050–4057, http://dx.doi.org/10.1016/j.apm.2010.03.036.
[17] P. Kumar, Typology of hydrologic predictability, Water Resour. Res. 47 (2011) W00H05, http://dx.doi.org/10.1029/2010WR009769.
[18] X. Wang, T. Lei, Hydrologic system behavior characteristic analysis and long-term prediction based on chaos radial basis function networks, BoletínTécnico 55 (2017) 536–546.
[19] W.A. Brock, W.D. Dechert, J.A. Scheinkman, B. LeBaron, A test for independence based on the correlation dimension, Econ. Rev. 15 (1996) 197–235.
[20] A. Kolmogorov, Logical basis for information theory and probability theory, IEEE Trans. Inform. Theory 14 (1968) 662–664, http://dx.doi.org/10.1109/TIT.1968.1054210.
[21] X.S. Zhang, R.J. Roy, E.W. Jensen, Eeg complexity as a measure of depth of anesthesia for patients, IEEE Trans. Biomed. Eng. 48 (2001) 1424–1433, http://dx.doi.org/10.1109/10.966601.
[22] J. Hu, J. Gao, J.C. Principe, Analysis of biomedical signals by the Lempel-Ziv complexity: the effect of finite data size, IEEE Trans. Biomed. Eng. 53 (2006) 2606–2609, http://dx.doi.org/10.1109/TBME.2006.883825.
[23] A. Lempel, J. Ziv, On the complexity of finite sequence, IEEE Trans. Inform. Theory 22 (1976) 75–81, http://dx.doi.org/10.1109/TIT.1976.1055501.
[24] D. Cerra, M. Datcu, Algorithmic relative complexity, Entropy 13 (2011) 902–914, http://dx.doi.org/10.3390/e13040902.
[25] N. Radhakrishnan, J.D. Wilson, P.C. Loizou, An alternate partitioning technique to quantify the regularity of complex time series, Int. J. Bifurc. Chaos 2000 (10) (2000) 1773–1779, http://dx.doi.org/10.1142/S0218127400001092.
[26] F. Kaspar, H.G. Schuster, Easily calculable measure for the complexity of spatiotemporal patterns, Phys. Rev. A 36 (1987) 842–848.
[27] D.T. Mihailović, M. Bessafi, S. Marković, I. Arsenić, S. Malinović-Milićević, P. Jeanty, M. Delsaut, J.-P. Chabriat, N. Drešković, A. Mihailović, Analysis of solar irradiation time series complexity and predictability by combining kolmogorov measures and hamming distance for la reunion (France), Entropy 20 (8) (2018) 570, http://dx.doi.org/10.3390/e20080570.
[28] E.G. Booth, J.F. Mount, J.H. Viers, Hydrologic variability of the cosumnes river floodplain, SFEWS4 (2) (2006) Article 2.
[29] G. Boeing, Visual analysis of nonlinear dynamical systems: chaos, fractals, self-similarity and the limits of prediction ,Systems 4 (4) (2016) 37, http://dx.doi.org/10.3390/systems4040037.
[30] M.T. Rosenstein, J.J. Collins, C.J. De Luca, A practical method for calculating largest lyapunov exponents from small data sets, Physica D 65 (1993) 117–134, http://dx.doi.org/10.1016/0167-2789(93)90009-P.
[31] M. Shapour, LYAPROSEN: MATLAB Function to Calculate Lyapunov Exponent, University of Tehran, Tehran, Iran, 2009.
[32] C. Rhodes, M. Morari, The false nearest neighbors algorithm: an overview, Comput. Chem. Eng. 21 (1997) S1149–S1154, http://dx.doi.org/10.1016/S0098-1354(97)87657-0.

[33] M. Lei, Z. Wang, Z. Feng, A method of embedding dimension estimation based on symplectic geometry, Phys. Lett. A 303 (2002) 179–189, http://dx.doi.org/10.1016/S0375-9601(02)01164-7.
[34] R. Huffaker, M. Bittelli, R. Rossa, Nonlinear Time Series Analysis with R, Oxford University Press, United Kingdom, 2017.
[35] T.W. Frison, H.D.I. Abarbanel, Ocean gravity waves: a nonlinear analysis of observations, J. Geophys. Res. 102 (C1) (1997) 1051–1059, http://dx.doi.org/10.1029/96JC02993.
[36] USGS, https://waterdata.usgs.gov/nwis, 2018 (accessed on 23 2018).
[37] C. Li, V.P. Singh, A.K. Mishra, Entropy theory-based criterion for hydrometric network evaluation and design: maximum information minimum redundancy, Water Resour. Res. 48 (5) (2012) 1–15.
[38] C. Heumann, M.S. Shalabh, Introduction to Statistics and Data Analysis, Springer International Publishing Switzerland, 2016.
[39] M. Tarnopolski, On the relationship between the hurst exponent, the ratio of the mean square successive difference to the variance, and the number of turning points, Phys. A 461 (2016) 662–673.
[40] R version 3.4.2, The R Foundation for Statistical Computing, 2017.
[41] J. Beran, Y. Feng, S. Ghosh, R. Kulik, Long-Memory Processes, Springer-Verlag Berlin Heidelberg, 2013.
[42] A. Berryman, M. Lima, M, Detecting the order of population dynamics from time series: nonlinearity causes spurious diagnosis, Ecology 88 (8) (2007) 2121–2123, http://dx.doi.org/10.1890/06-0609.1.
[43] L.S. McCue, A.W. Troesch, Use of Lyapunov exponents to predict chaotic vessel motions, in: Proceedings: Seventh International Ship Stability Workshop, Shangai, China, 2004.
[44] A. Adewumi, J. Kagamba, A. Alochukwu, Application of chaos theory in the prediction of motorized traffic flows on urban networks, Math. Probl. Eng. (2016) 5656734, http://dx.doi.org/10.1155/2016/5656734, 15.
[45] A.R. Tănasie, G.R. Drăcea, A. Lădaru, Chaos theory perspective on international migration, Sustainability 9 (2017) 2346–2355, http://dx.doi.org/10.3390/su9122355.
[46] H.R. Wang, C. Wang, X. Lin, J. Kang, An improved arima model for hydrological simulations, Nonlin. Processes Geophys. Discuss. 21 (2014) 841–876, http://dx.doi.org/10.5194/npgd-1-841-2014.
[47] M. Tarnopolski, Correlation between the hurst exponent and the maximal Lyapunov exponent: examining some low-dimensional conservative maps, Phys. A 490 (2018) 834–844.
[48] J. Gao, J. Hu, W. Tung, W.-W, Facilitating joint chaos and fractal analysis of biosignals through nonlinear adaptive filtering, PLoS ONE 6 (2011) E24331, http://dx.doi.org/10.1371/journal.pone.0024331.